Structural Causal Bandits with non-manipulable variables

Sanghack Lee Elias Bareinboim

Purdue University

Executive Summary

- **SCM-MAB** = MAB (problem) + Causality (principle)

- Studied **structural properties** of SCM-MAB

  (**MIS**) some arms share the same reward

  (**POMIS**) some arms are worth playing

  (**z^{2}ID**) express one arm's reward w/ other arms samples

- **SCM-MAB algo** = MAB algo + structural properties

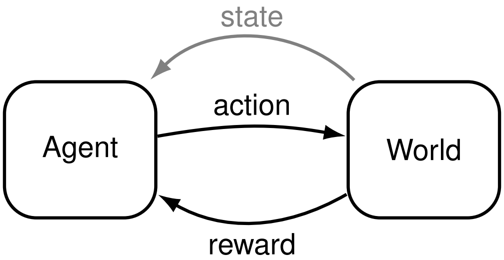
- Better performance due to

  $\rightarrow$ a smaller # of qualified arms

  $\rightarrow$ more accurate estimation

Motivation

AI Agent



Reinforcement Learning	World is **stateful**
Multi-Armed Bandit*	World is **stateless**

Multi-Armed Bandit

A *classic*, sequential decision-making problem

Given a set of K **arms** (= **actions**), $\mathbf{A}$

arms' reward distributions, $\{\nu_i\}_{1 \le i \le K}$

$(\mu_k \doteq \mathbb{E}_{Y \sim \nu_k}[Y], \qquad \mu^* \doteq \max_{k \in \mathbf{A}} \mu_k)$

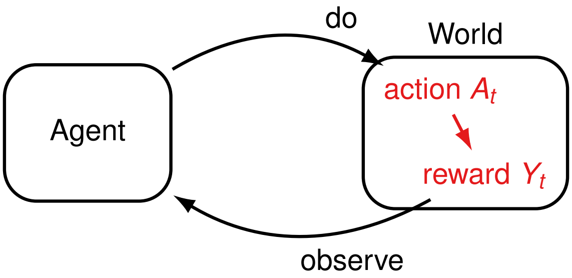
Play at every round t, an agent plays an arm A_t, and get a stochastic **reward** $Y_t \sim \nu_t$.

Goal to minimize **cumulative regret** in T rounds

$$\mathrm{Reg}_T \doteq T\mu^* - \sum_{t=1}^{T} \mathbb{E}[Y_t]$$

Challenge a trade-off between **exploitation** vs. **exploration**

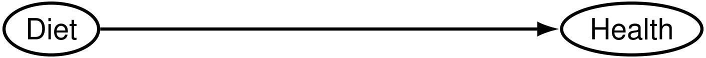
AI Agent (again)



Graphical Understanding of MAB

$P(Y \mid A)$

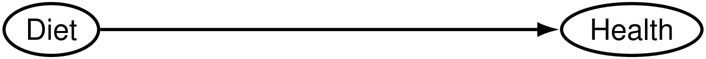
Clinical Trials



1. **MAB** is at the highest level of **abstraction.**

2. **MAB** is (often) all about **intervention.**

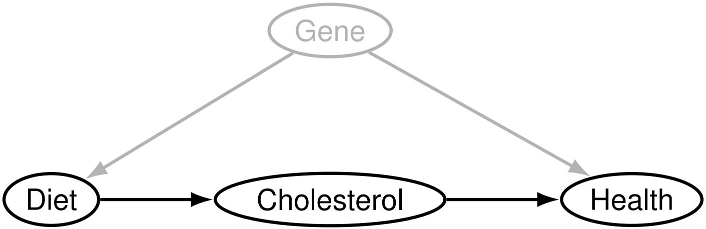
Clinical Trials



1. **MAB** is at the highest level of **abstraction.**
   $\rightarrow$ Where are other variables?

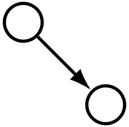
2. **MAB** is (often) all about **intervention.**
   $\rightarrow$ People may choose their own diets.

Clinical Trials



1. **Causal Structure —** *less* abstract, *more* informative.
   → Physicians can observe some variables.

2. **Passive Observation**
   → Physicians know that people choose their own diets.



How can we utilize causal knowledge in solving MAB problems?

SCM-MAB

— MAB on SCM

SCM — *the* Causal Framework

Definition (Structural Causal Model)

SCM $\mathcal{M}$ is a 4-tuple $\langle \mathbf{V}, \mathbf{U}, P(\mathbf{U}), \mathbf{F} \rangle$

$\mathbf{V}$	observed variables
$\mathbf{U}$	unobserved variables
$P(\mathbf{U})$	a joint distribution of $\mathbf{U}$
$\mathbf{F}$	a set of functions for $\mathbf{V}$

SCM — *the* Causal Framework

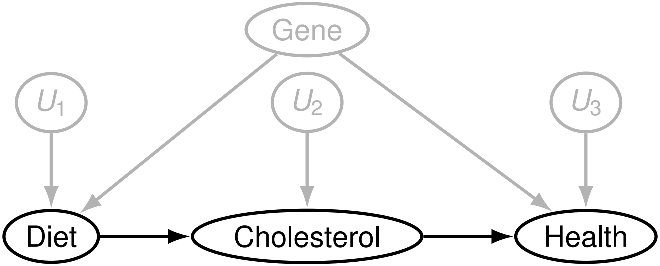
Definition (Structural Causal Model)

SCM $\mathcal{M}$ is a 4-tuple $\langle \mathbf{V}, \mathbf{U}, P(\mathbf{U}), \mathbf{F} \rangle$

$\mathbf{V}$	observed variables
$\mathbf{U}$	unobserved variables
$P(\mathbf{U})$	a joint distribution of $\mathbf{U}$
$\mathbf{F}$	a set of functions for $\mathbf{V}$

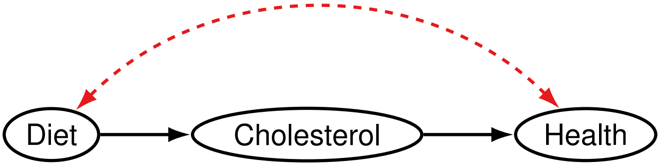
- induces a **causal graph** $\mathcal{G}$
- defines **interventional distributions**, e.g., $P(Y \mid do(\mathbf{x}))$ (a passive observation as an *empty* intervention.)

SCM: Example & Causal Graph



- $\mathbf{V} = \{\text{Diet}, \text{Cholesterol}, \text{Health}\}$
- $\mathbf{U} = \{U_1, U_2, U_3, \text{Gene}\}$
- $\mathbf{F} = f_{\text{Diet}}(U_1, \text{Gene}),\ f_{\text{Chol}}(U_2, \text{Diet}),\ f_{\text{Health}}(\text{Gene}, U_3, \text{Chol})$
- $P(\mathbf{U}) = P(U_1, U_2, U_3, \text{Gene})$

SCM: Example & Causal Graph



- Unobserved Confounders (UCs) as bidirected edges.
- **U** other than UCs are not shown.

SCM: Example & Causal Graph

- Here, *do*(*diet*) deletes the bidirected edge.
- Health is still affected by Gene.

SCM-MAB

Definition (SCM-MAB)

A tuple of a SCM $\mathcal{M}$ and a reward variable $Y \in \mathbf{V}$.

SCM-MAB

Definition (SCM-MAB)

A tuple of a SCM $\mathcal{M}$ and a reward variable $Y \in \mathbf{V}$.

SCM-MAB induces:

Intervention Sets	all subsets of $\mathbf{V}$ except Y i.e., $2^{\mathbf{V} \setminus \{Y\}}$
Arms	all possible values for intervention sets i.e., $\mathbf{A} = \{\mathbf{x} \in \mathfrak{X}_{\mathbf{X}} \mid \mathbf{X} \in 2^{\mathbf{V} \setminus \{Y\}}\}$
Reward	$\nu_{\mathbf{x}} = P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$ $\mu_{\mathbf{x}} = \mathbb{E}[Y \mid do(\mathbf{x})]$

SCM-MAB

Definition (SCM-MAB)

A tuple of a SCM $\mathcal{M}$ and a reward variable $Y \in \mathbf{V}$.

SCM-MAB induces:

Intervention Sets	all subsets of $\mathbf{V}$ except Y $\{\emptyset, \{\text{Diet}\}, \{\text{Chol}\}, \{\text{Diet}, \text{Chol}\}\}$
Arms	all possible values for intervention sets $\{\textit{diet}{:}\textit{vegan}\}, \{\textit{diet}{:}\textit{poke}, \textit{chol}{:}\textit{low}\}, \ldots$
Reward	$\nu_{\mathbf{x}} = P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$ $\mu_{\mathbf{x}} = \mathbb{E}[Y \mid do(\mathbf{x})]$

SCM-MAB w/ Non-manipulability

Definition (SCM-MAB)

A tuple of a SCM $\mathcal{M}$ and a reward variable $Y \in \mathbf{V}$
with non-manipulable variables $\mathbf{N} \subset \mathbf{V} \setminus \{Y\}$

SCM-MAB w/ **N** induces:

Intervention Sets	all subsets of **V** except **N** and Y i.e., $2^{\mathbf{V} \setminus \mathbf{N} \setminus \{Y\}}$
Arms	all possible values for intervention sets i.e., $\mathbf{A} = \{\mathbf{x} \in \mathfrak{X}_{\mathbf{X}} \mid \mathbf{X} \in 2^{\mathbf{V} \setminus \mathbf{N} \setminus \{Y\}}\}$
Reward	$\nu_{\mathbf{x}} = P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$ $\mu_{\mathbf{x}} = \mathbb{E}[Y \mid do(\mathbf{x})]$

SCM-MAB w/ Non-manipulability

Definition (SCM-MAB)

A tuple of a SCM $\mathcal{M}$ and a reward variable $Y \in \mathbf{V}$
with non-manipulable variables $\mathbf{N} \subset \mathbf{V} \setminus \{Y\}$

SCM-MAB w/ $\mathbf{N} = \{\text{Cholesterol}\}$ induces:

Intervention Sets all subsets of $\mathbf{V}$ except $\mathbf{N}$ and Y
i.e., $\{\emptyset, \{\text{Diet}\}\}$

Arms all possible values for intervention sets
$\{\}, \{\textit{diet}:\textit{vegan}\}, \ldots$

Reward $\nu_{\mathbf{x}} = P(Y \mid do(\mathbf{x})) = P_{\mathbf{x}}(Y)$
$\mu_{\mathbf{x}} = \mathbb{E}[Y \mid do(\mathbf{x})]$

SCM-MAB w/ Non-manipulability

Setting $\langle \mathcal{M}, Y, \mathbf{N} \rangle$
(arms, reward distributions, etc are all induced)

Goal to minimize a cumulative regret (same as MAB)

Assumption 1. can access to the causal graph $\mathcal{G}$
$\rightarrow$ an agent sees $\langle \mathcal{G}, Y, \mathbf{N} \rangle$

2. can observe $\mathbf{v}$ after each play (not just y)

SCM-MAB w/ Non-manipulability

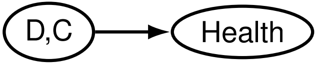
Setting $\langle \mathcal{M}, Y, \mathbf{N} \rangle$
(arms, reward distributions, etc are all induced)

Goal to minimize a cumulative regret (same as MAB)

Assumption 1. can access to the causal graph $\mathcal{G}$
$\rightarrow$ an agent sees $\langle \mathcal{G}, Y, \mathbf{N} \rangle$

2. can observe $\mathbf{v}$ after each play (not just y)

existing MAB algorithms work!



w/ **explicit ‘no-op’**

How can we utilize the causal graph $\mathcal{G}$ and observations $\mathbf{v}$?

How can we utilize the causal graph $\mathcal{G}$ and observations $\mathbf{v}$?

What are some properties of SCM-MAB to be exploited?

Structural Properties of SCM-MAB

Structural Properties in SCM-MAB

A **traditional MAB** assumes that arms are independent**.**

In **SCM-MAB**, arms are dependent due to the shared causal mechanism.

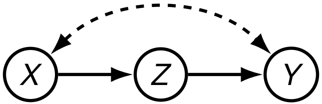
Structural Properties in SCM-MAB

A **traditional MAB** assumes that arms are independent.

In **SCM-MAB**, arms are dependent due to the shared causal mechanism.

1. Equivalence two arms share the **same** reward distribution
2. Partial-orders an intervention set is $\geq$ to the other set w.r.t. their **best achievable expected rewards**
3. Expressions inferring one arm's reward distribution from other arms' samples.

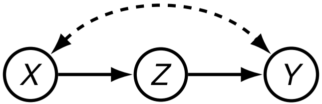
Structural Property 1: Equivalence



$$\mu_{X,Z} = \mu_Z$$

$\because (Y \perp\!\!\!\perp X \mid Z)_{\mathcal{G}_{\overline{X,Z}}}$, Rule 3 of *do*-calculus (Pearl, 2000)

Structural Property 1: Equivalence

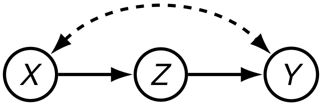


$$\mu_{X,Z} = \mu_Z$$

$\because (Y \perp\!\!\!\perp X \mid Z)_{\mathcal{G}_{\overline{X,Z}}}$, Rule 3 of *do*-calculus (Pearl, 2000)

Implication: prefer playing $do(Z)$ to playing $do(X, Z)$

Structural Property 1: Equivalence



$$\mu_{x,z} = \mu_z$$

$\because (Y \perp\!\!\!\perp X \mid Z)_{\mathcal{G}_{\overline{X,Z}}}$, Rule 3 of *do*-calculus (Pearl, 2000)

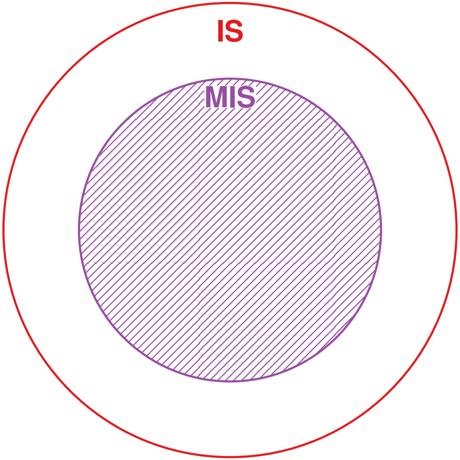
Implication: prefer playing $do(Z)$ to playing $do(X, Z)$

Definition (Minimal Intervention Set, MIS)

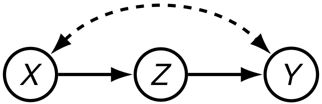
(informal) the smallest IS among ISs sharing the same reward

Structural Property 1: Equivalence

MISs $\subseteq$ ISs



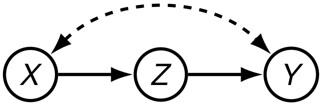
Structural Property 2: Partial-orderedness



$$\mu_x = \sum_z \mu_z P(z|x) \leq \sum_z \mu_{z^*} P(z|x) = \mu_{z^*}$$

where $\mu_{z^*} = \max_{z \in \mathfrak{X}_Z} \mu_z$ (the best achievable reward by $do(Z)$)

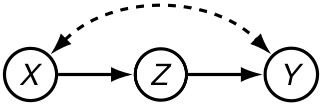
Structural Property 2: Partial-orderedness



$$\mu_{X^*} \leq \mu_{Z^*}$$

Implication: playing *do*(Z) is preferred to playing *do*(X).

Structural Property 2: Partial-orderedness



$$\mu_{x^*} \leq \mu_{z^*}$$

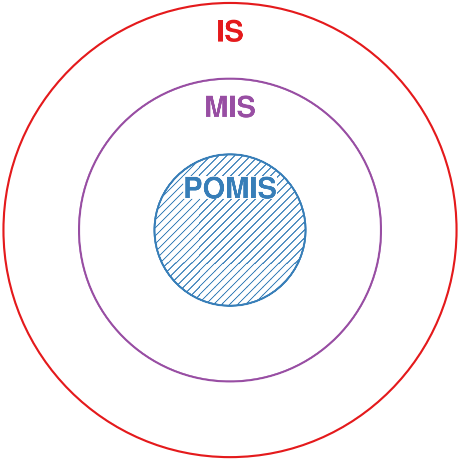
Implication: playing $do(Z)$ is preferred to playing $do(X)$.

Definition (Possibly-Optimal MIS, POMIS)

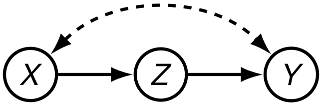
(informal) an MIS that can be optimal in some model conforming to the causal graph.

Structural Property 2: Partial-orderedness

POMISs $\subseteq$ MISs $\subseteq$ ISs



Structural Properties 1 & 2



$\mathbf{N} = \emptyset$

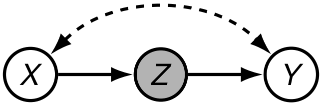
ISs $\emptyset, \{Z\}, \{X\}, \{X, Z\}$

MISs $\emptyset, \{Z\}, \{X\}$

POMISs $\emptyset, \{Z\}$

See LB (2018) for the characterization of POMIS when $\mathbf{N} = \emptyset$.

Structural Properties 1 & 2



$\mathbf{N} = \{Z\}$

ISs $\emptyset, \{X\}$

MISs $\emptyset, \{X\}$

POMISs $\emptyset, \{X\}$

Applying POMIS algorithm (for $\mathbf{N} = \emptyset$) on the **Latent Projection** of $\mathcal{G}$ over $\mathbf{V} \setminus \mathbf{N}$ yields valid POMISs under $\mathbf{N} \neq \emptyset$.

Structural Property 3: Expressions Relating Arms

- A **traditional MAB** assumes that arms are independent.
  Playing an arm $\mathbf{x}$ informs nothing about arm $\mathbf{x}'$.

- In **SCM-MAB**, arms are dependent.
  Playing an arm $\mathbf{x}$ informs something about arm $\mathbf{x}'$.

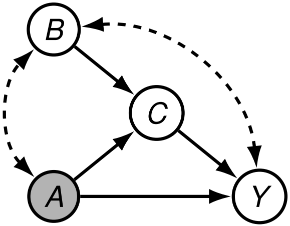
Structural Property 3: Expressions Relating Arms

- A **traditional MAB** assumes that arms are independent.
  Playing an arm $\mathbf{x}$ informs nothing about arm $\mathbf{x}'$.

- In **SCM-MAB**, arms are dependent.
  Playing an arm $\mathbf{x}$ informs something about arm $\mathbf{x}'$.

  We proposed $\mathbf{z^2}$**ID** algorithm — represents a query with available quantities

Structural Property 3: Expressions Relating Arms — an example



POMISs are $\emptyset$, $\{B\}$, and $\{C\}$.

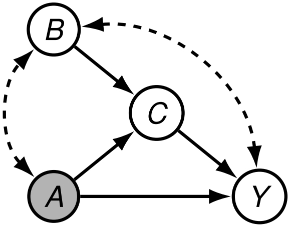
$$P(y) = \sum_{a,b,c} P_b(c|a) P_c(a, b, y)$$

$$P_b(y) = \sum_{a,c} P(c|a, b) \sum_{b'} P(y|a, b', c) P(a, b')$$

$$P_c(y) = \sum_{a,b} P(y|a, b, c) P(a, b)$$

$$P_c(y) = \sum_{a} P_b(y|a, c) P_b(a)$$

Structural Property 3: Expressions Relating Arms — an example



POMISs are $\emptyset$, $\{B\}$, and $\{C\}$.

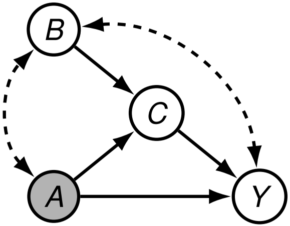
$$P(y) = \sum_{a,b,c} P_b(c|a) P_c(a,b,y)$$

$$P_b(y) = \sum_{a,c} P(c|a,b) \sum_{b'} P(y|a,b',c) P(a,b')$$

$$P_c(y) = \sum_{a,b} P(y|a,b,c) P(a,b)$$

$$P_c(y) = \sum_{a} P_b(y|a,c) P_b(a)$$

Structural Property 3: Expressions Relating Arms — an example



POMISs are $\emptyset$, $\{B\}$, and $\{C\}$.

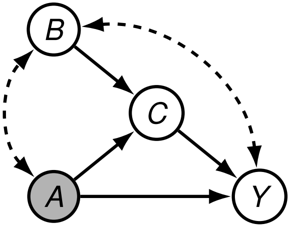
$$P(y) = \sum_{a,b,c} P_b(c|a) P_c(a, b, y)$$

$$P_b(y) = \sum_{a,c} P(c|a, b) \sum_{b'} P(y|a, b', c) P(a, b')$$

$$P_c(y) = \sum_{a,b} P(y|a, b, c) P(a, b)$$

$$P_c(y) = \sum_{a} P_b(y|a, c) P_b(a)$$

Structural Property 3: Expressions Relating Arms — an example



POMISs are $\emptyset$, $\{B\}$, and $\{C\}$.

$$P(y) = \sum_{a,b,c} P_b(c|a) P_c(a, b, y)$$

$$P_b(y) = \sum_{a,c} P(c|a, b) \sum_{b'} P(y|a, b', c) P(a, b')$$

$$P_c(y) = \sum_{a,b} P(y|a, b, c) P(a, b)$$

$$P_c(y) = \sum_{a} P_b(y|a, c) P_b(a)$$

SCM-MAB algorithms

Incorporating Structural Properties into MAB algos.

What we know,

- POMIS all arms vs. possibly-optimal arms
- z^2ID utilize samples from other (POMIS) arms

Incorporating Structural Properties into MAB algos.

What we know,

POMIS all arms vs. possibly-optimal arms

z^2ID utilize samples from other (POMIS) arms

2 algorithms we considered:

TS **posterior distributions** for expected rewards

kl-UCB **upper confidence bounds** for expected rewards

Incorporating Structural Properties into MAB algos.

What we know,

POMIS	all arms vs. possibly-optimal arms
z^2ID	utilize samples from other (POMIS) arms

2 algorithms we considered:

z^2-TS	**posterior distributions** for expected rewards $\rightarrow$ **adjust** ‘posterior distributions’ reflecting all used data.
z^2-kl-UCB	**upper confidence bounds** for expected rewards $\rightarrow$ **adjust** ‘upper bounds’ by taking account samples from other arms.

SCM-MAB algorithm: modified TS

taking advantage of **POMIS** and **z²ID**.

function z²-TS($\mathcal{G}, Y, \mathbf{N}, T$)

$\mathbb{Z} \leftarrow \mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y}$

$\mathbf{A} \leftarrow \{\mathbf{x} \in \mathfrak{X}_{\mathbf{X}} \mid \mathbf{X} \in \mathbb{Z}\}$

$\hat{\boldsymbol{\theta}}_{\mathbf{x}} \leftarrow \{P_{\mathbf{x}}(y)\} \cup \{\text{z}^2\text{ID}(\mathcal{G}, y, \mathbf{x}, \mathbb{Z}')\}_{\mathbb{Z}' \subseteq \mathbb{Z} \setminus \{\mathbf{X}\}}$ **for** $\mathbf{x} \in \mathbf{A}$

$\mathbf{D} \leftarrow \{D_{\mathbf{x}} = \emptyset\}_{\mathbf{x} \in \mathbf{A}}$

for t in $1, \ldots, T$ **do**

  for $\mathbf{x} \in \mathbf{A}$ **do**

    $\hat{\theta}_{\mathbf{x}}, \hat{s}^2_{\mathbf{x}} \leftarrow \text{bMVWA}(\mathbf{D}, \hat{\boldsymbol{\theta}}_{\mathbf{x}})$

    Find $\hat{\alpha}_{\mathbf{x}}, \hat{\beta}_{\mathbf{x}}$ such that $\text{Beta}(\hat{\alpha}_{\mathbf{x}}, \hat{\beta}_{\mathbf{x}})$ matching $\hat{\theta}_{\mathbf{x}}, \hat{s}^2_{\mathbf{x}}$

    $\theta_{\mathbf{x}} \sim \text{Beta}(\hat{\alpha}_{\mathbf{x}}, \hat{\beta}_{\mathbf{x}})$

  $\mathbf{x}' \leftarrow \arg\max_{\mathbf{x} \in \mathbf{A}} \theta_{\mathbf{x}}$

  Sample $\mathbf{v}$ by $do(\mathbf{x}')$ and append $\mathbf{v}$ to $D_{\mathbf{x}'}$

SCM-MAB algorithm: modified kl-UCB

taking advantage of **POMIS** and **z²ID**.

function z^2-KL-UCB$(\mathcal{G}, Y, \mathbf{N}, T, f \leftarrow \ln(t) + 3\ln(\ln(t)))$

Initialize $\mathbb{Z}, \mathbf{A}$, $\{\hat{\boldsymbol{\theta}}_{\mathbf{x}}\}_{\mathbf{x}\in\mathbf{A}}$, $\mathbf{D}$

$(\forall_{\mathbf{x}\in\mathbf{A}})$ Sample $\mathbf{v}$ by $do(\mathbf{x})$, and append $\mathbf{v}$ to $D_{\mathbf{x}}$

for t in $|\mathbf{A}|, \ldots, T$ **do**

$\hat{\theta}_{\mathbf{x}}, \hat{s}^2_{\mathbf{x}} \leftarrow$ bMVWA$(\mathbf{D}, \hat{\boldsymbol{\theta}}_{\mathbf{x}})$ **for** $\mathbf{x} \in \mathbf{A}$

$\hat{N}_{\mathbf{x}} \leftarrow \hat{\theta}_{\mathbf{x}}(1 - \hat{\theta}_{\mathbf{x}})/\hat{s}^2_{\mathbf{x}}$; $\quad \hat{t} \leftarrow \sum_{\mathbf{x}} \hat{N}_{\mathbf{x}}$

$$\boldsymbol{\mu} = \left\{ \sup \left\{ \mu \in [0,1] : \mathrm{KL}(\hat{\theta}_{\mathbf{x}}, \mu) \leq \frac{f(\hat{t})}{\hat{N}_{\mathbf{x}}} \right\} \right\}_{\mathbf{x}\in\mathbf{A}}$$

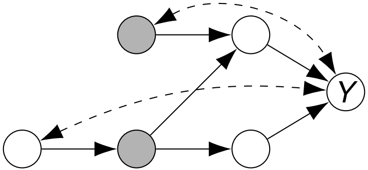
$\mathbf{x}' \leftarrow \arg\max_{\mathbf{x}\in\mathbf{A}} \mu_{\mathbf{x}}$

Sample $\mathbf{v}$ by $do(\mathbf{x}')$, and append $\mathbf{v}$ to $D_{\mathbf{x}'}$

Empirical Evaluation

Experimental settings

- **4** strategies: Brute-force (all ISs), MIS, POMIS, POMIS+
- **2** base MAB algorithms: TS, kl-UCB
- **3** SCM-MAB problems, e.g.,

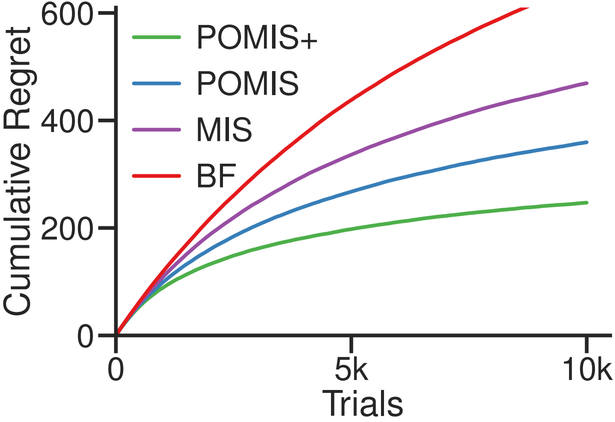


- **1000** simulations

Experimental results

Performance: POMIS+ $>$ POMIS $\geq$ MIS $\geq$ Brute-force

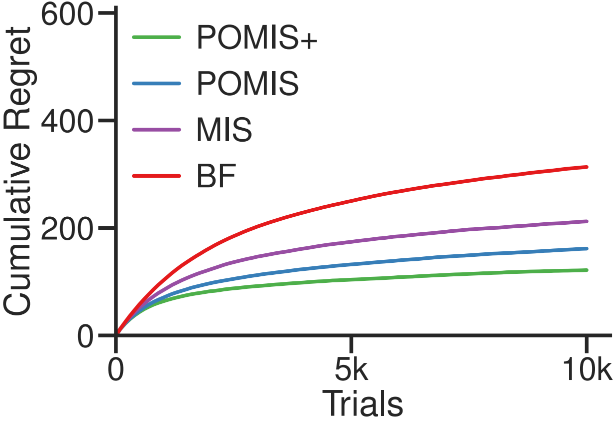
base algorithm: **kl-UCB**



Experimental results

Performance: POMIS+ $>$ POMIS $\geq$ MIS $\geq$ Brute-force

base algorithm: **TS**



Conclusions

Conclusions

How can we make better decision w/ causal knowledge?

$\because$ Causal mechanisms *do* exist.

$\because$ There are *tools* for causal inference.

$\because$ Ignoring causal mechanisms might behave suboptimally.

We

defined **SCM-MAB** w/ non-manipulability constraints

studied **3 structural properties** of SCM-MAB

devised SCM-MAB **algorithms** w/ the structural properties

Conclusions

How can we make better decision w/ causal knowledge?

∵ Causal mechanisms *do* exist.

∵ There are *tools* for causal inference.

∵ Ignoring causal mechanisms might behave suboptimally.

We

defined **SCM-MAB** w/ non-manipulability constraints

studied **3 structural properties** of SCM-MAB

devised SCM-MAB **algorithms** w/ the structural properties

Mahalo!

(thank you)